███████████

FEBRUARY 1$^{\text{ST}}$, 2016

(1) Warm-ups

(A) Which bands come to ACL Fest?

Is it true that if a band plays at Lollapalooza, then it is more likely to play at Austin City Limits (ACL) that year? To be able to provide an answer to this question, we should look at the relative risk of attending ACL given that a band has attended Lollapalooza that year. The counts used are displayed below in Table 1.

Table 1: Table representing the counts of bands that attended Lollapalooza between 2008 and 2011.

	Lollapalooza	
	Did Not Attend ($n = 800$)	Attended ($n = 438$)
Did Not Attend ACL	719	361
Attended ACL	81	77

Here, we see that the relative risk of a band attending ACL for bands that attended Lollapalooza is

$$\text{Relative risk} = \frac{77/(77+361)}{81/(81+719)} = \frac{77/438}{81/800} = 1.736$$

indicating that a band that attends Lollapalooza is 1.736 times more likely to attend ACL that year than a band that does not attend Lollapalooza.

Table 2 contains that corresponding attendance counts for other festival pairings involving ACL.

Table 2: Table representing the counts of bands that attended at least one of Bonnaroo, Coachella, and Outside Lands between 2008 and 2011, with numbers in parentheses indicating the total number of bands in the corresponding column. (N = Did Not Attend, Y = Attended)

	Bonnaroo		Coachella		Outside Lands	
	N (916)	Y (322)	N (686)	Y (552)	N (1046)	Y (192)
Did Not Attend ACL	817	263	590	490	914	166
Attended ACL	99	59	96	62	132	26

The relative risks for attending ACL are 1.695 for bands attending Bonnaroo, 0.803 for bands attending Coachella, and 1.073 for bands attending Outside Lands. From this, it seems that if you want to see your favorite band members at ACL, you should hope that they do not attend Coachella! In contrast, attending Bonnaroo or Outside Lands (or Lollapalooza as mentioned above) increases the chance of being able to see the band at ACL. Note that this analysis of the data does not tell us about interactions; for example, maybe band attendance at a specific combination of Bonnaroo, Coachella, Lollapalooza, and Outside Lands in one year is associated with the greatest chance of attending ACL.

(B) How many calories do people eat at Chipotle?

In Figure 1, we have created a histogram of the data of calorie counts of 3,042 meals ordered from Chipotle through GrubHub with bin widths equal to 40 calories. Noticeable spikes at 700, 900, 1,000, 1,200, 1,400, and 1,600 calories are likely present due to the fact that some meals are more frequently ordered than others, as well as the possibililty that the typical classes of meals available at Chipotle are comprised of meals that all have similar calorie counts to one another.

The mean number of calories of Chipotle meals is approximately 1,093.674. The number of meals with more than 1,600 calories is 343. A central 80% coverage interval of this distribution would be (620, 1,614.5).
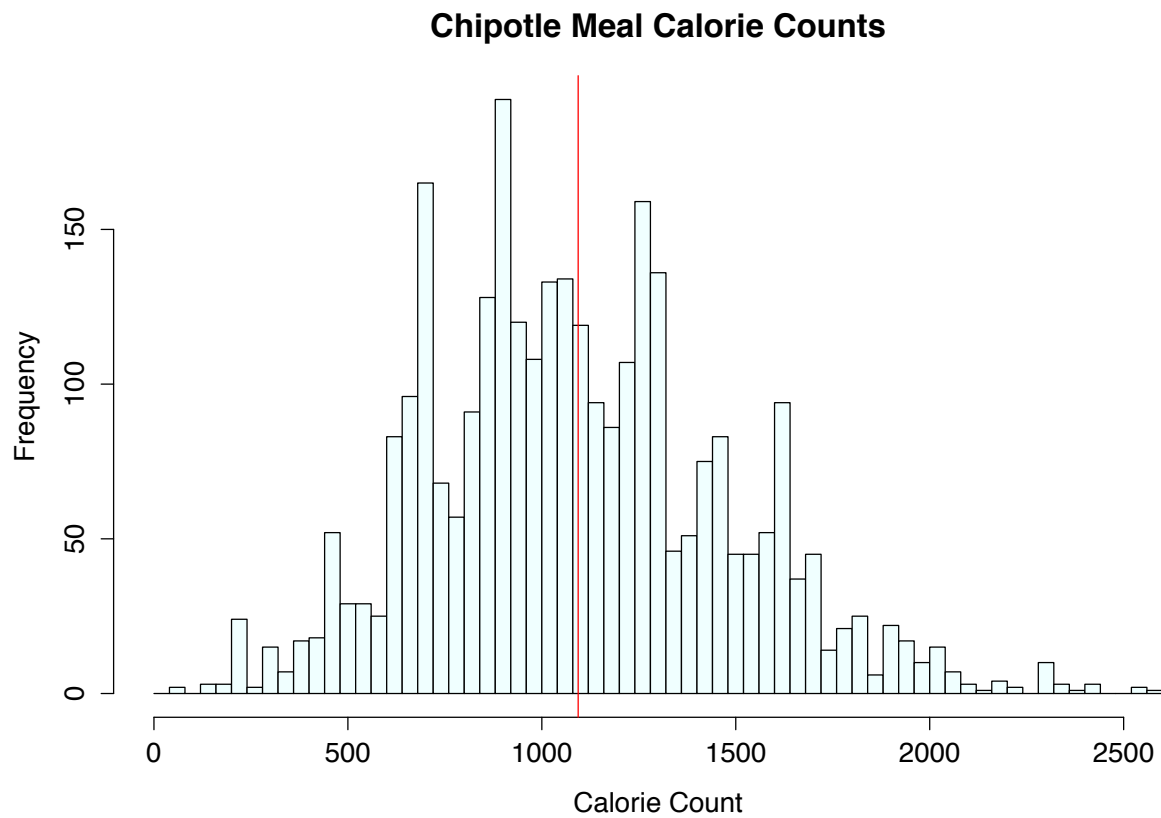
Chipotle Meal Calorie Counts

Figure 1: Histogram of the calorie counts of 3,042 meals ordered from Chipotle through GrubHub. The mean of 1,093.674 calories is indicated by the red vertical line overlaid on the histogram.

(2) Exploring multivariate data

Intuition tells us that newer cars are likely to have higher asking prices attached to them than older cars. An interesting question to ask would be whether the trim of a car modifies this general relationship. Figure 2 shows us price versus model year stratified by trim for a data set of various Mercedes S-Class vehicles. One observation we might make is that it is clear that 450 is not a particularly prevalent trim option (for reasons seemingly unrelated to price, as the one S-Class car with 450 trim was selling for less than $50,000). Additionally, it appears that the 430, 500, and 55 AMG options were discontinued around 2006 to be supplanted by the 550, 63 AMG, and 65 AMG options.

Though the general trend of newer cars being more expensive than older cars is reflected in half of the plots, this trend is not readily visible in the plots for the 320, 420, 430, 450, 500, and 55 AMG options, which have prices that appear largely constant regardless of model year (age must be taken into account here). In addition, though we might expect the depreciation of the value of S-Class cars to resemble exponential decay (using a model dependent on car age), it appears that the 63 AMG and 65 AMG trim options defy this in favor of a linear progression, suggesting that they may have a better ability to retain value over time. The 600 trim option appears to most resemble the expected exponential model. The 550 trim option also appears to follow this trajectory, but with much higher variation in pricing for all of its active model years (2007-2015). The plots also reflect the growth of the Mercedes S-Class line as the number of distinct data points for each model year generally increases over time.

In summary, our lattice plot tells us about much more than changing pricing in the Mercedes S-Class line. Within the plot lies a story about the recent history of the S-Class line, with the rise and fall of trim options as cultural tastes and managerial decisions change with the progression of time, as well as the overall growth in popularity of S-Class cars.
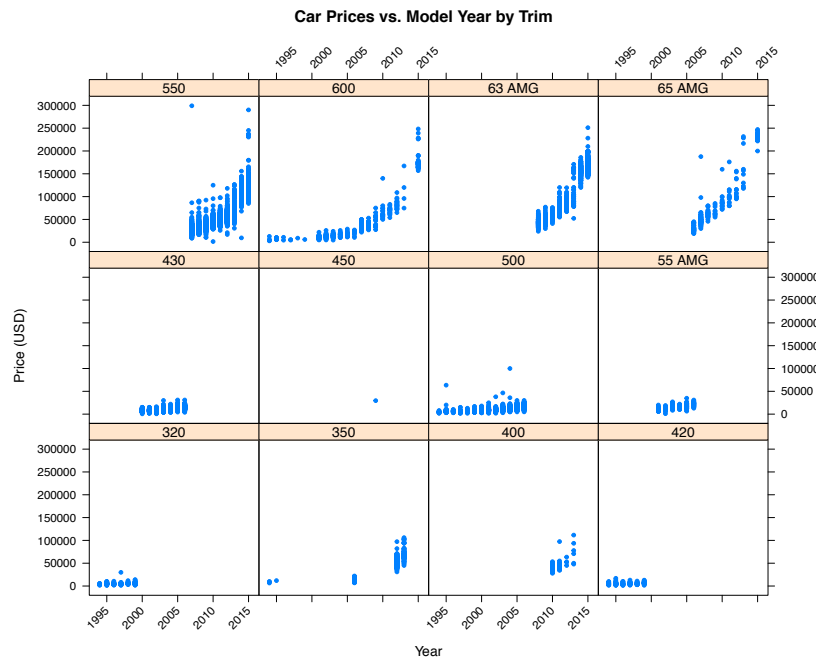

Figure 2: Lattice plot of the relationship between price, model year, and color of over 29,000 Mercedes S-Class vehicles advertised on the secondary automobile market in 2014.

(3) Austin food critics

Upon visual inspection of boxplots of meal prices of restaurants in Central Austin neighborhoods (Figure 3), we can see that the Congress Avenue area can be considered the most expensive neighborhood in terms of restaurant meals due to the high median meal price of $80, with Second Street coming in, well, second with a median meal price of $60. On the other end of the spectrum, the restaurants of the Drag have the lowest median meal price at $15, with Bouldin Creek following as the second least expensive neighborhood for restaurant food with a median meal price of $20. Interestingly, comparing mean meal prices instead of median meal prices still leads to Congress Avenue/Second Street and the Drag/Bouldin Creek being considered the two most expensive and two least expensive neighborhoods, respectively.
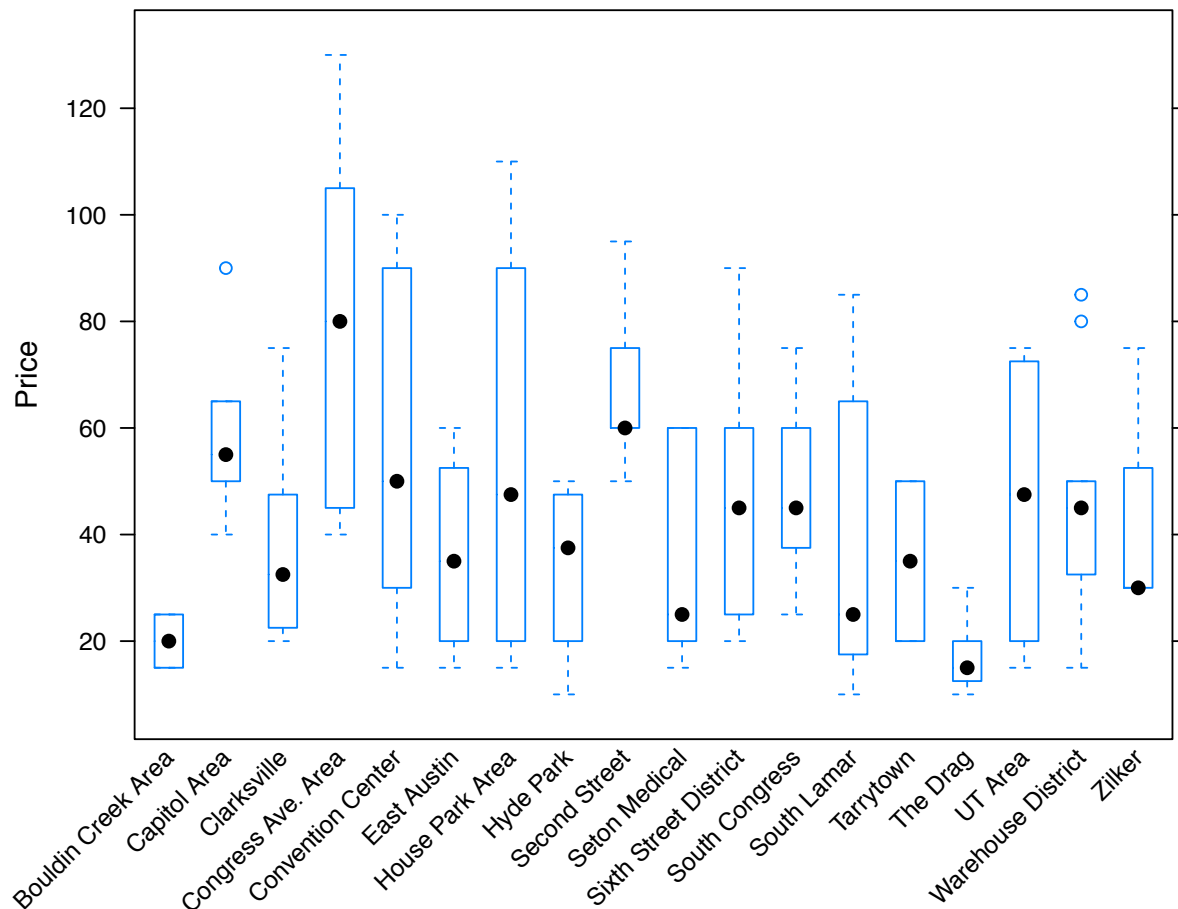

Figure 3: Boxplots of meal prices of restaurants in Central Austin neighborhoods.

Comparing correlation coefficients, we find that food quality is a better predictor of the price of a meal than restaurant atmosphere (compare $r = 0.525$ with $r = 0.219$). Figure 4 shows two scatterplots, each with superimposed trendlines.
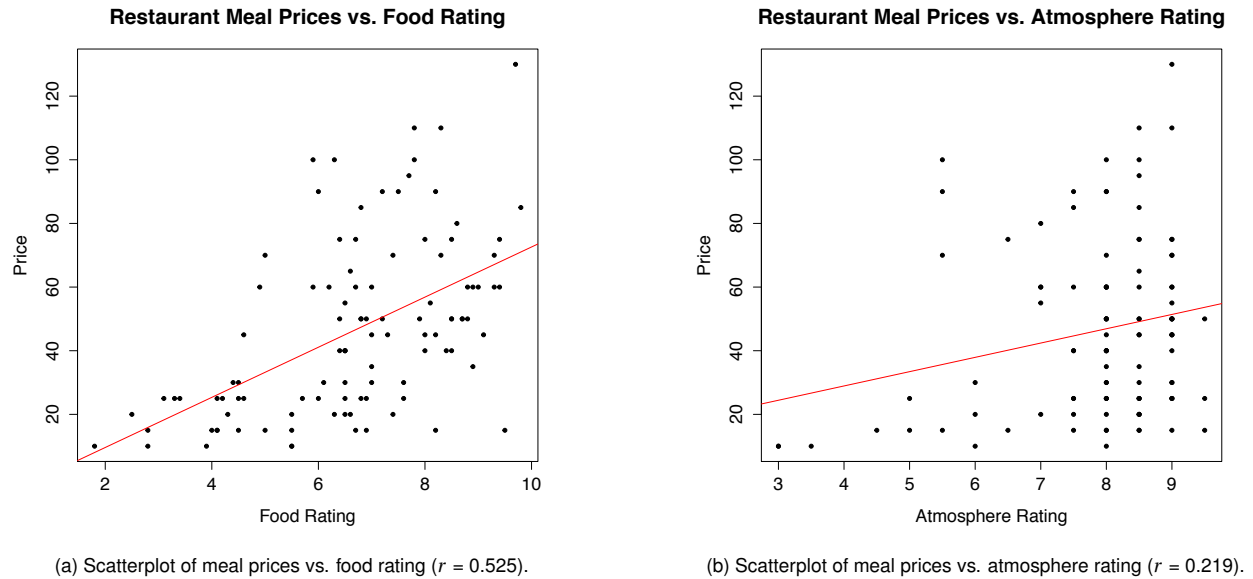
Restaurant Meal Prices vs. Food Rating

(a) Scatterplot of meal prices vs. food rating ($r = 0.525$).

Restaurant Meal Prices vs. Atmosphere Rating

(b) Scatterplot of meal prices vs. atmosphere rating ($r = 0.219$).

Figure 4: Scatterplot of meal prices vs. ratings.

To compute a "food-adjusted value" measure for each restaurants, we just use the differences between our predicted prices from our model based on food rating and the observed prices (*i.e.,* the residuals). Essentially, the best-value restaurants are the ones that exceed our expectations (in a positive way) in terms of how much we actually end up paying for a meal. To compare the neighborhoods, we construct a boxplot and look at the median values of the residuals by neighborhood; Figure 5 shows us that under our definition of value, The Drag and Hyde Park are the "best-value" neighborhoods, while the Convention Center area and Congress Avenue area have the "worst-value" restaurants. Additionally, we can construct a scatterplot to compare restaurants on an individual level; this is done in Figure 6.

Restaurant Price Residuals

Figure 5: Boxplot of food-adjusted values (residuals) by neighborhood. Lower residuals indicate better values due to the actual cost of meals being lower than our predicted values.
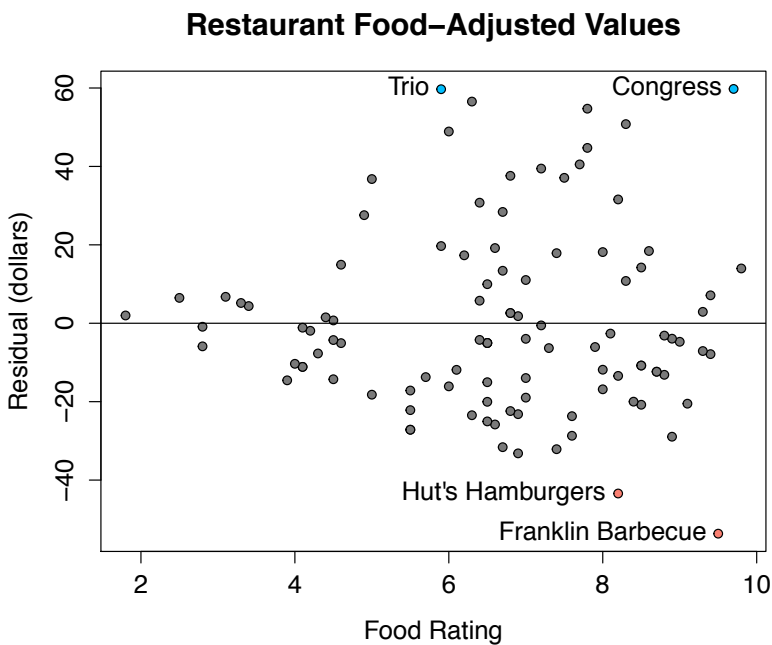
Restaurant Food–Adjusted Values

Figure 6: Scatterplot of food-adjusted values (residuals) vs. food rating. Franklin Barbecue and Huf's Hamburgers are the "best-value" restaurants in Central Austin, while Congress and Trio are the "worst-value" restaurants.